

Structural compliance: A new metric for protein flexibility

Original

Structural compliance: A new metric for protein flexibility / Scaramozzino, D.; Khade, P. M.; Jernigan, R. L.; Lacidogna, G.; Carpinteri, A.. - In: PROTEINS. - ISSN 0887-3585. - STAMPA. - 88:11(2020), pp. 1482-1492. [10.1002/prot.25968]

Availability:

This version is available at: 11583/2847748 since: 2020-10-06T20:11:42Z

Publisher:

John Wiley and Sons Inc.

Published

DOI:10.1002/prot.25968

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

Structural Compliance – A New Metric for Protein Flexibility

Domenico Scaramozzino^a, Pranav M. Khade^b, Robert L. Jernigan^{*b}, Giuseppe Lacidogna^a,
Alberto Carpinteri^a

^a*Department of Structural, Geotechnical and Building Engineering, Politecnico di Torino, Corso
Duca degli Abruzzi 24, 10129, Torino, Italy*

^b*Roy J. Carver Department of Biochemistry, Biophysics and Molecular Biology, Iowa State
University, Ames, IA, 50011, USA*

*Corresponding author: jernigan@iastate.edu

Acknowledgments

This research has been supported by NSF grant DBI-1661391 and by NIH grants R01-GM127701 and R01-GM127701-01S1. We also thank Research IT @Iowa State University for helping with some aspects of the computing.

Abstract

Proteins are the active players in performing essential molecular activities throughout biology, and their dynamics has been broadly demonstrated to relate to their mechanisms. The intrinsic fluctuations have often been used to represent their dynamics and then compared to the experimental B-factors. However, proteins do not move in a vacuum and their motions are modulated by solvent which can impose forces on the structure. In this paper, we introduce a new structural concept, which has been called the structural compliance, for the evaluation of the global and local deformability of the protein structure in response to intramolecular and solvent forces. Based on the application of pairwise pulling forces to a protein elastic network, this structural quantity has been computed and sometimes is even found to yield an improved correlation with the experimental B-factors, meaning that it may serve as a better metric for protein flexibility. The inverse of structural compliance, namely the structural stiffness, has also been defined, which shows a clear anti-correlation with the experimental data. Although the present applications are made to proteins, this approach can also be applied to other biomolecular structures such as RNA. This present study considers only elastic network models, but the approach could be applied further to conventional atomic molecular dynamics. Compliance is found to have slightly better agreement with the experimental B-factors, perhaps reflecting its bias toward the effects of local perturbations, in contrast to mean square fluctuations. The code for calculating protein compliance and stiffness is freely accessible at bit.ly/PACKMAN-compliance.

Keywords: Elastic Network Model, Protein Flexibility, Structural Compliance, Protein Stiffness, B-factor.

Introduction

Elastic network models have been used for several decades by researchers to understand the features and mechanisms of proteins and macromolecules. Since the seminal work of Tirion¹, models exploiting simple harmonic potentials have been found to provide significant information about the equilibrium fluctuations of the proteins, without the need to include detailed potentials and run computationally-expensive Molecular Dynamics (MD) simulations.

Among the different elastic models previously studied, one of the most popular has been the coarse-grained Anisotropic Network Models (ANM), firstly developed by Atilgan et al.², which treats the protein as a simple network of nodes connected by linear springs, all having the same force constant. Also, it has been shown that even coarse-grained models at the residue-level, i.e. considering only C^α atoms, are able to accurately predict the protein dynamics^{2,3}. Surprisingly, atomic and coarse-grained models yield nearly identical results⁴⁻⁶. This is largely because the overall shape of the structure is the most important property for its dynamics. This model, although very simple, generally has been effective in describing the directions of protein motions in their thermal fluctuations. Furthermore, the low-frequency motions arising from ANM calculations were found to agree well with the conformational changes shown by some proteins upon ligand-binding^{7,8}. Several variations of the original ANM were developed, e.g. taking into account distance-dependent spring constants^{9,10}, considering groups of residues as rigid blocks^{11,12}, using finite-element-based approaches¹³, etc.

Besides the evaluation of the low-frequency modes and thermal fluctuations, the ANM was also useful for understanding the pathway and the nature of conformational changes. For example, Kim et al.¹⁴ used elastic network models in order to generate realistic, feasible transition pathways for a protein between two conformations. With the aim of predicting the end conformation starting from one known structure of the protein, Atilgan et al.^{15,16} developed a perturbation-response scanning (PRS) method, where localized forces were applied to the network at specific positions and the response of the structure was evaluated and compared to the conformational transition. Following the same linear-response approach, more recently, Liu et al.¹⁷ made use of the ANM to investigate the conformational change of the GroEL subunit and the directionality of the force applied from the ATP hydrolysis site.

The usual procedure to validate the accuracy of an ANM-like elastic network has been based on the evaluation of the computed fluctuations which are calculated from the ANM normal modes¹⁸, and compare these to the experimental B-factors which are extracted from crystallographic experiments and reported into the PDB file¹⁹. Experimental B-factors result from the local uncertainty about the position of a certain atom and, although they can result from a combination of different factors, for very high-resolution structures they can provide information about the inherent flexibility of the protein. Other comparisons have been made between sets of experimental structures and the computed normal modes^{20,21}. Khade et al.²² recently used the experimental B-factors to verify a packing-based method for the prediction of the flexible and rigid parts of proteins.

In Structural Mechanics, the flexibility of a structure can be assessed in several ways. One of these is based on the concept of structural compliance, which can be defined as the displacement in a certain direction per unit force. According to this definition, it is clear that compliance is the inverse of structural stiffness. Since we know that experimental B-factors should account for the local and global flexibility, the idea behind the present work was to use the concept of structural

compliance as a measure of a protein's inherent flexibility. For this purpose, we model a large number of proteins by means of the pfANM¹⁰, we then evaluate the distribution of the structural compliance along the protein chain, based on a pairwise force application methodology (see Methods), and compare the resulting distribution with the experimental B-factors. In order to judge the efficacy of the structural compliance for predicting the protein flexibility, we also evaluate the theoretical fluctuations from the pfANM normal modes and compare them against the experimental B-factors. From the results, we conclude that structural compliance can be effectively used as a novel and, in some cases, better measure to describe the protein flexibility. Note that the methodology we use for deriving the compliance is substantially different from the concept recently introduced by Arikawa²³ to study the softer motions.

In addition to providing some evaluation of the intrinsic flexibility of a protein, the concept of compliance here is also presumed to simulate effects of Brownian motions resulting from random collisions between protein and solvent molecules. The ability of a protein to switch from one conformation to another is central to its function. It is therefore essential to study the protein dynamics under the influence of surrounding water and any other intermolecular interactions. Because proteins are usually surrounded by water and other small molecules and ions, they undergo constant bombardment by these molecules. However, there are no effective models to account for these effects on protein dynamics, except for Brownian dynamics. **Other possible approaches to deal with the effect of external solvent rely on Langevin dynamics²⁴.** We know that a particle undergoes Brownian motion if it is in a medium containing discrete particles such as water molecules²⁵. Because of this, many studies have tested²⁶⁻²⁸ and modelled protein motions as Brownian motions in Brownian dynamics simulations. Along with the Brownian motion resulted from solvent, intramolecular interactions contribute to details of the overall motion. Often these intramolecular interactions and/or random collisions can affect kinetics in a protein^{16,29,30} making it important to properly model these motions. The concept of structural compliance considered here essentially relies on stretching two atoms of the protein apart as if this motion were resulting from the collision/repulsion with solvent (besides intramolecular interactions). This points towards the possibility of using compliance to model the effects of the Brownian motions resulting from random collisions with solvent. This is fundamentally different from the intrinsic fluctuations of a protein, **calculated by the traditional elastic network models**, where the dynamics do not include any effects from solvent. As we will show, these two quantities exhibit subtle differences between them.

The software for the calculation of compliance and stiffness values has been inserted into the PACKMAN package, developed by Khade et al.²², which is an open source Python-based package and can be freely accessed at bit.ly/PACKMAN-compliance .

Methods

In this Section, we recount the fundamentals of the Anisotropic Network Model (ANM), which allows the evaluation of the intrinsic fluctuations based on normal modes, and then provide the framework for the definition and calculation of the structural compliance.

Anisotropic Network Model (ANM)

The main idea behind the ANM is that the most important features of proteins and macromolecules, such as their fluctuations and global mechanisms, can be derived making use of

simple elastic networks, made up of nodes connected by linear springs, without the need to include different interaction potentials depending on specific atom or residue types.

For a system of N points, e.g. N residues in the C^α -atom only representation, the Hessian matrix of the structure takes the following form²

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_{1,1} & \dots & \mathbf{H}_{1,i} & \dots & \mathbf{H}_{1,j} & \dots & \mathbf{H}_{1,N} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \mathbf{H}_{i,1} & \dots & \mathbf{H}_{i,i} & \dots & \mathbf{H}_{i,j} & \dots & \mathbf{H}_{i,N} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \mathbf{H}_{j,1} & \dots & \mathbf{H}_{j,i} & \dots & \mathbf{H}_{j,j} & \dots & \mathbf{H}_{j,N} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \mathbf{H}_{N,1} & \dots & \mathbf{H}_{N,i} & \dots & \mathbf{H}_{N,j} & \dots & \mathbf{H}_{N,N} \end{bmatrix}, \quad (1)$$

where each 3×3 $\mathbf{H}_{i,j}$ submatrix contains the interaction information between nodes i and j . Each of the submatrices can be computed based on the harmonic potential associated with the elastic spring connecting nodes i and j , which can be written as

$$V_{i,j} = \frac{\gamma_{i,j}}{2} (r_{i,j} - r_{i,j}^0)^2, \quad (2)$$

where $\gamma_{i,j}$ stands for the force constant of the spring connecting nodes i and j , $r_{i,j}^0$ is the initial equilibrium distance between the two points, with $r_{i,j}$ being their separation distance after deformation. By computing the second derivatives of $V_{i,j}$ with respect to the three Cartesian XYZ directions, one can then express the off-diagonal $\mathbf{H}_{i,j}$ submatrix as

$$\mathbf{H}_{i,j} = \begin{bmatrix} \frac{\partial^2 V_{i,j}}{\partial x_i \partial x_j} & \frac{\partial^2 V_{i,j}}{\partial x_i \partial y_j} & \frac{\partial^2 V_{i,j}}{\partial x_i \partial z_j} \\ \frac{\partial^2 V_{i,j}}{\partial y_i \partial x_j} & \frac{\partial^2 V_{i,j}}{\partial y_i \partial y_j} & \frac{\partial^2 V_{i,j}}{\partial y_i \partial z_j} \\ \frac{\partial^2 V_{i,j}}{\partial z_i \partial x_j} & \frac{\partial^2 V_{i,j}}{\partial z_i \partial y_j} & \frac{\partial^2 V_{i,j}}{\partial z_i \partial z_j} \end{bmatrix} = -\gamma_{i,j} \begin{bmatrix} \frac{(x_j - x_i)^2}{r_{i,j}^0} & \frac{(x_j - x_i)(y_j - y_i)}{r_{i,j}^0} & \frac{(x_j - x_i)(z_j - z_i)}{r_{i,j}^0} \\ \frac{(y_j - y_i)(x_j - x_i)}{r_{i,j}^0} & \frac{(y_j - y_i)^2}{r_{i,j}^0} & \frac{(y_j - y_i)(z_j - z_i)}{r_{i,j}^0} \\ \frac{(z_j - z_i)(x_j - x_i)}{r_{i,j}^0} & \frac{(z_j - z_i)(y_j - y_i)}{r_{i,j}^0} & \frac{(z_j - z_i)^2}{r_{i,j}^0} \end{bmatrix}. \quad (3)$$

Finally the diagonal terms $\mathbf{H}_{i,i}$ can be calculated as a summation involving all the nodes connected to node i , according to the following equation

$$\mathbf{H}_{i,i} = - \sum_{j=1, j \neq i}^N \mathbf{H}_{i,j}. \quad (4)$$

It is clear that the evaluation of the Hessian matrix is dependent on some model parameters, which are the geometrical cut-off and the distribution of the force constants among the springs of the network⁹. The original ANM was developed by ignoring the interaction specificity between different residues and their separation distance in the three-dimensional space, so that equal spring constants were assigned between all close members of the network and a geometrical cut-off was

applied in order to consider springs placed only between nodes having an initial distance below this chosen parameter⁹. Later on, distance-dependent force constants were introduced, as

$$\gamma_{i,j} \propto \frac{1}{r_{i,j}^{0p}}, \quad (5)$$

p being the parameter which reduces the spring constants for longer interaction distances. It was shown that non-zero p values enable obtaining better results for predicting the protein fluctuations^{9,10}. The concept of an inverse power dependence has also been exploited in order to remove the need to consider an explicit geometrical cut-off limit (pfANM)¹⁰. Here, we use the pfANM, with a value of $p = 3$, which we found yields the best results. A complete analysis was performed by considering different values of the exponent p , i.e. 1, 2, 3, 4, 6, 12, and those results are available in the Supplementary Material.

Calculation of the computed fluctuations

Once the pfANM Hessian matrix is computed based on the protein coordinates obtained from the PDB file, the standard eigenvalue-eigenvector decomposition is carried out in order to evaluate the $3N$ eigenvalues and $3N$ eigenvectors. Since the protein structure is not externally constrained, the first six eigenvalues are all zero and the corresponding mode shapes are associated with the 6 rigid body rotation and translation motions. The modal decomposition takes the following form

$$\mathbf{H} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T, \quad (6)$$

where \mathbf{U} represents the matrix containing the $3N-6$ non-trivial eigenvectors and $\mathbf{\Lambda}$ is the diagonal matrix containing the $3N-6$ non-zero eigenvalues λ_n . Each column n of \mathbf{U} represents the $3N \times 1$ vector δ_n , containing the displacements of the N residues according to the mode shape n . For each residue i , the $3i-2$, $3i-1$ and $3i$ entries of vector δ_n reflect the node's mode displacements along X, Y and Z direction, respectively. Once the dynamical features are extracted from the Hessian matrix, one can easily calculate the computed fluctuations as¹⁸

$$B_i = \frac{8}{3} \pi^2 k_B T \sum_{n=7}^{3N} \frac{\delta_{i,n}^2}{\lambda_n}, \quad (7)$$

where B_i represents the computed B-factor for residue i , k_B is the Boltzmann constant, T is the absolute temperature and $\delta_{i,n}$ stands for the total absolute displacement of node i according to the eigenvector n . $\delta_{i,n}$ is evaluated as the square root of the sum of the squares of the displacements along the X, Y and Z direction. Finally, the Pearson correlation coefficient can be calculated between the distributions of the computed fluctuations from Eq. (7) and the experimental B-factors, in order to assess the accuracy of the elastic network model.

Calculation of the structural compliance

Compliance is a measure of the flexibility of a structure in a certain direction, when subjected to specific loading conditions, and it is usually measured in units of meters per newton (m/N). For instance, consider a simple unidimensional spring, having a force constant γ . If we apply a couplet of forces F along the spring direction, the process will result in an elongation of the spring $\delta = F/\gamma$ (see Fig. 1a). The structural compliance of the system is then defined as the total resulting displacement in the direction of the force divided by the absolute value of the force, i.e. $C = \delta/F = 1/\gamma$. As can be seen, in this elementary case, the compliance can be simply expressed as the inverse of the spring stiffness.

Here, we use this structural concept in order to obtain insight into the flexibility of a protein, modeled as an elastic network. Specifically, the value of the structural compliance for each couplet of residues i and j is first calculated by applying equal and opposite forces to each pair of residues i and j (Fig. 1b). The unit force vectors \mathbf{F}_i and \mathbf{F}_j are defined so that their orientation corresponds to the direction of \mathbf{R}_{ij}^0 , being $\mathbf{R}_{ij}^0 = \{x_j - x_i, y_j - y_i, z_j - z_i\}^T$ the distance vector separating residues i and j in the initial configuration. The three components of the force vectors can then be simply calculated from

$$\mathbf{F}_i = \frac{-\mathbf{R}_{ij}^0}{\sqrt{(x_j - x_i)^2 + (y_j - y_i)^2 + (z_j - z_i)^2}}, \mathbf{F}_j = \frac{\mathbf{R}_{ij}^0}{\sqrt{(x_j - x_i)^2 + (y_j - y_i)^2 + (z_j - z_i)^2}}, \quad (8)$$

where x_i , y_i and z_i are the crystallographic coordinates of residue i , and x_j , y_j and z_j the coordinates of residue j . The force components in Eq. (8) are then inserted into the global force vector \mathbf{F} ($3N \times 1$), which contains all the forces acting upon each node of the network. Note that, since we are pulling only two residues at a time, the global force vector \mathbf{F} will only have six non-zero components, corresponding to the six values reported in Eq. (8).

Defining $\boldsymbol{\delta}$ as the $3N \times 1$ displacement vector, which contains the displacements of the nodes, and under the assumption of linear elastic behavior and small displacements, we can express the Hooke's generalized law for this elastic network as

$$\mathbf{F} = \mathbf{H}\boldsymbol{\delta}. \quad (9)$$

Defining the force vector \mathbf{F} from Eq. (8), and computing the Hessian matrix from Eqs. (3) and (4), the displacements of the structure, when subjected to the couplets of forces, are straightforwardly computed from

$$\boldsymbol{\delta} = \tilde{\mathbf{H}}^{-1}\mathbf{F}, \quad (10)$$

where $\tilde{\mathbf{H}}^{-1}$ denotes the pseudo-inverse of the Hessian matrix. This is calculated from the $3N-6$ eigenvalues λ_n and eigenvectors \mathbf{U}_n

$$\tilde{\mathbf{H}}^{-1} = \sum_{n=7}^{3N} \frac{\mathbf{U}_n \mathbf{U}_n^T}{\lambda_n}. \quad (11)$$

The pairwise structural compliance $C_{i,j}$, referring to residues i and j ($i, j = 1 \dots N, i \neq j$), can finally be evaluated by calculating the total distance variation between the residues along the pulling direction, i.e. the total directional displacement, divided by the magnitude of the force acting on the nodes. Note that, since the forces have been defined as unit vectors, the structural compliance basically corresponds to the total displacement along the pulling direction. Therefore, once the displacements of residues i (δ_i) and j (δ_j) are **selected** from the global displacement vector δ , $C_{i,j}$ can be simply computed as

$$C_{i,j} = \frac{(\delta_j - \delta_i)^T \mathbf{F}_j}{|\mathbf{F}_j|} = \frac{(\delta_j - \delta_i)^T \mathbf{R}_{ij}^0}{\sqrt{(x_j - x_i)^2 + (y_j - y_i)^2 + (z_j - z_i)^2}}. \quad (12)$$

By iterating the procedure over each pair of residues i and j , ignoring $i = j$, a compliance map is generated for the entire protein, which provides significant details about the rigidity and flexibility of the structure. If we calculate the inverse of $C_{i,j}$, we can obtain a measure of the stiffness along the i - j direction, i.e. $S_{i,j} = 1/C_{i,j}$, and consequently obtain the stiffness map of the structure as well.

Finally, we define the total compliance for each residue i , i.e. C_i , as the average value of all the pairwise compliances involving residue i , i.e.

$$C_i = \frac{1}{N-1} \sum_{j \neq i} C_{i,j}. \quad (13)$$

Similarly, given the pairwise stiffness values $S_{i,j}$ for each residue pair, the global stiffness values can be calculated by averaging all the pairwise stiffnesses

$$S_i = \frac{1}{N-1} \sum_{j \neq i} S_{i,j}. \quad (14)$$

The Pearson correlation coefficient can then be calculated between the compliance/stiffness profiles from Eqs. (13) and (14) and the experimental B-factors distribution. **Note that, although the simplest one, the standard averaging procedure from Eqs. (13) and (14) is not the only one possible choice, e.g. one might employ a weighted average depending on the distance between residues i and j , etc. Obviously, different averaging strategies might cause slight differences in the compliance/stiffness profile and, ultimately, in the prediction of protein flexibility, which can be the objective for further optimization purposes.**

The approach presented here to evaluate the pairwise stiffness $S_{i,j}$ is **quite different** to the one used by Eyal and Bahar³¹, **although it provides a similar metric**. In that paper, the authors showed that the mechanical resistance maps derived from elastic network models were effective to predict the anisotropic response of proteins when subjected to external forces, e.g. during single-molecule manipulation experiments. However, in that case, the mechanical resistance in the i - j direction was

not computed based on the direct structural definition, as in the present work, but was rather based on the ANM normal modes.

Protein datasets

The correlation analyses were performed on two different datasets, which are referred to as Dataset 1 and Dataset 2. Both include X-ray protein structures downloaded from the PDB¹⁹. Dataset 1 includes only single-chain proteins, whereas Dataset 2 contains only multi-chain proteins. Table 1 gives the characteristics of the high-quality PDB structures in each set.

Results and Discussion

In Fig. 2a and 2b the results are reported for the calculations performed on Dataset 1, for the single-chain proteins. Specifically, Fig. 2a shows the distributions of the correlation coefficients obtained with the experimental B-factors when considering the normal-mode-based fluctuations (grey histogram) or the structural compliance (black histogram). The distribution of the Pearson correlation coefficients with the experimental B-factors are strongly similar for both the fluctuations and the structural compliance. We observe that the median value (M) of the distributions referring to the structural compliance (0.66) is slightly higher than the one associated with the ANM fluctuations (0.63), and that the standard deviation (σ) for the compliance is slightly lower. This confirms that the structural compliance introduced here is useful as a measure of protein flexibility.

Besides compliance, the other structural concept which was introduced above and was found to provide useful insights about the protein flexibility, is the structural stiffness, which can be calculated according Eq. (14). The correlation between the stiffness and experimental B-factors has been tested for all the 921 single-chain proteins in Dataset 1 and the results are shown in Fig. 2b. The stiffness and the experimental B-factors are strongly anti-correlated, with the median value equal to -0.64. These results confirm that the stiffness defined here can be an effective additional metric to characterize protein rigidity and flexibility. The results from the complete analysis, varying the values of p , are reported into the Supplementary Material in Figs. S1 and S2.

In some individual cases, we also observe that the compliance is more strongly correlated with the experimental B-factors than the computed fluctuations. For example, by considering the X-ray structure of the Human Complement Protein C8 γ (PDB: 1LF7) and taking into account the elastic network model with $p = 3$, a correlation coefficient of 0.61 is obtained when considering the fluctuations (Fig. 3a), whereas a value of 0.79 is found for the compliance (Fig. 3b). As can clearly be seen in Fig. 3b, the profile of the structural compliance, calculated according Eq. (13), shows good agreement with the experimental B-factors, and therefore it allows the identification of the flexible and rigid portions of the protein. Figure 3b also compares the experimental B-factors for Human Complement Protein C8 γ with the stiffness values. A Pearson coefficient of -0.76 was obtained for this comparison. It can be seen that the two profiles are strongly anti-correlated, meaning that when the experimental B-factors show upward peaks (high-flexibility regions), the stiffness profile shows downward peaks (low-stiffness regions). This allows us to conclude that the structural stiffness defined here can provide insights on the internal protein flexibility. Note that the anti-correlation value of the stiffness and the correlation for the compliance are very similar, i.e. -0.76 and 0.79, but they are not identical. In fact, although the pairwise stiffness $S_{i,j}$ is defined as the

inverse of the pairwise compliance $C_{i,j}$, the strict inverse relationship is lost in the total S_i and C_i values, because of the averaging procedure reported in Eqs. (13) and (14). For this reason, the stiffness and compliance profiles are nearly, but not exactly, mirrored (see Fig. 3b). In Fig. 3c, the normalized values presented in Fig. 3a and 3b are mapped onto the protein structure. As can be seen, although there are similarities between compliances and fluctuations, compliances agree qualitatively more closely with the experimental B-factors.

As specified above, in order to calculate the profiles of the structural compliance, it is first necessary to generate a map of pairwise compliances, which provides information about the distribution of the deformability over the structure. In Fig. S3, the complete compliance and stiffness maps for the elastic network model of Human Complement Protein C8 γ are reported. See the Supplementary Material for more details.

As indicated above, Dataset 2 has a smaller number of proteins with multiple chains. Considering the two separate datasets was carried out to assess whether the concepts of structural compliance and stiffness can provide insights on the flexibility of proteins, regardless the number of amino acid chains.

Similarly to the results presented in Fig. 2a, Fig. 4a shows the outcomes of the correlation coefficients for the fluctuations and compliance when compared to the experimental B-factors for Dataset 2. The results are almost identical to what has been already obtained for Dataset 1. This means that the concept of structural compliance is indeed useful also for predicting the flexibility of multi-chain macromolecules. In the same way, Fig. 4b shows the distributions of the correlation coefficients which were obtained when comparing the stiffness profiles to the experimental B-factors for Dataset 2. Again, a good anti-correlation was found, showing that the stiffness concept can be also used for multi-chain proteins as a measure of their inherent flexibility. The results from the complete analysis, by varying the values of p , are reported into the Supplementary Material in Figs. S4 and S5.

An example is shown in Fig. 5 for the profiles of the fluctuations, compliance and stiffness compared to the experimental B-factors of the dimeric protein *Clitocybe nebularis* ricin B-like lectin (PDB: 3NBC). The correlation with the computed fluctuations is found to be 0.73; whereas the correlations with the structural compliance and stiffness are 0.80 and -0.78, respectively. As can be seen from Fig. 5b, the peaks corresponding to the high-flexibility regions found in the experimental B-factors distribution are well described by the compliance profile. In the same way, a good anti-correlation is found with the stiffness profile. In Fig. 5c, the normalized values are mapped onto the structure of the dimeric protein. In this case the compliance is more able to quantify the flexibility on the external tips of the protein as well as its deformability in the cores. In the Supplementary Material, in Fig. S6, the complete normalized compliance and stiffness maps for the same protein are reported, which show a clear symmetry for the two chains of the molecule.

Another useful consideration is related to the concept of packing. Protein packing is a multiscale phenomenon which influences the local and global dynamics of the protein. It is often informative to study which parts of a protein are so densely packed as to be completely immobile and which are less dense, where internal movements are possible, to make conclusions about the mechanism and function of the protein²². The compliance and stiffness maps proposed in this paper can serve as another way to study protein packing. The compliance and stiffness profiles can help us to determine the dynamic communities³²⁻³⁴ and to coarse-grain them for further analysis. **We can use a hinge prediction method (PACKMAN²²) based on packing densities to identify the hinges in the parts of the protein that tend to be more flexible, making these sites responsible for the global**

deformability. So, if these regions are predicted to be hinges and show high compliance values, they can most likely represent global deformability, whereas local deformability occurs otherwise.

As shown in Figs. 2 and 4, the distributions of the correlation coefficients obtained when comparing the structural compliance or fluctuations to the experimental B-factors were very similar. That means that, on average, when considering a large number of protein structures, almost the same correlation coefficient could be obtained when comparing the compliance or the fluctuations profiles to the experimental data. However, as already shown in Figs. 3 and 5, for various cases an enhanced agreement with the experimental data was found when calculating the structural compliance, rather than considering the computed fluctuations. In some other cases, the correlation between the structural compliance and the experimental B-factors is found to be lower than the one obtained with the computed fluctuations. This outcome can be understood by considering the plots in Fig. 6, which shows the statistical distribution of the differences Δ between the correlation coefficients for Dataset 1 (Fig. 6a) and Dataset 2 (Fig. 6b). For each protein structure, the value of Δ is defined as

$$\Delta = \rho_{C,B_{exp}} - \rho_{F,B_{exp}}, \quad (15)$$

where $\rho_{C,B_{exp}}$ is the correlation coefficient from comparing the compliance to the experimental B-factors, and $\rho_{F,B_{exp}}$ the same quantity related to the fluctuations. It is then clear that positive Δ values imply higher correlations with the structural compliance, whereas negative Δ values indicate higher correlations with the fluctuations.

It can be seen from Figs. 6a and 6b that the histogram of Δ values resembles a zero-centered Gaussian distribution, although a certain right-oriented skewness is recognizable toward positive Δ values. Therefore, we cannot conclude that the structural compliance improves the correlation with experimental data for all cases, but this is true at least for more than half of the protein set. At this point, one could try and understand why, for certain cases, the experimental B-factors show an improved agreement with the structural compliance while, for some other cases, they are more correlated to the normal-mode-based fluctuations. Although one might argue that this should depend on certain protein features, e.g. the overall size, the globularity, etc., it was not possible to identify any correlation.

One possible explanation could be related to the very definition of the fluctuations and structural compliance introduced here. As can be seen from Eq. (7), the fluctuations are evaluated based on the eigenvectors derived from the ANM Hessian matrix, weighting their contribution according to the corresponding eigenvalues. In this way, the low-frequency modes, which are usually the larger-scale global motions of the protein, reflecting the global deformability, play a major role in defining the fluctuations. Therefore, the local deformability is not usually well-reflected into the computed fluctuations, since the corresponding motions usually occur at higher frequencies and, for this reason, they are underweighted. Contrarily, when calculating the structural compliance according to Eqs. (12) and (13), both the local and global flexibility of the molecule are probed equally. As a matter of fact, when applying pairwise pulling forces on close residues the local flexibility of the protein is probed, whereas when the forces act on a distant pair of amino acids the global deformability of the structure is assessed for the most part. Therefore, the different correlation coefficients obtained when comparing the fluctuations or the structural compliance

against the experimental B-factors might just be related to the different amount of local vs global flexibility that the experimental data reflect.

Moreover, it must be also noted that the experimental B-factors are not error-free and might not necessarily account for the actual relative balance between global-local flexibility of the protein structure, as they may contain some unavoidable errors from the crystallography experiment. Nevertheless, properties extracted from these simple models, like the fluctuations and structural compliance, can still provide significant information about the protein flexibility.

Conclusions

In this paper a new Structural Mechanics-based concept was introduced in order to investigate protein flexibility, which has been referred to as the structural compliance. The calculation of this new quantity was based on application of pairwise pulling forces at the all pairs of nodes in an elastic network model, in order to probe both the global and local flexibility of the protein. This concept has been used here since it may also simulate some of the effects of Brownian motions resulting from random collisions with the external environment. Interestingly, the profiles of the structural compliance along the protein chain have been found to show good correlation with the experimental B-factors derived from crystallographic experiments. The results have been confirmed by investigating large datasets of single- and multi-chain protein structures. Also, for various cases, it has been found that the structural compliance shows a slightly enhanced agreement with experimental data, higher than the one obtained when considering the classical ANM-based fluctuations. Therefore, it can be concluded that this structural quantity can be effectively used as a new measure for predicting the protein inherent flexibility. Also, it suggests that the ANM fluctuations do not always adequately represent the importance of the local fluctuations. Another structural concept was also introduced, which is strictly related to the structural compliance, and which has been defined as the structural stiffness. Likewise, a good anti-correlation has been found when comparing the experimental B-factors against the stiffness profiles, showing that regions with high B-factors are likely to show low stiffness values.

Both of the structural concepts introduced here not only allow the investigation of the flexibility of protein structures, but they can also lead to two-dimensional representations, showing the pairwise residue-residue deformability/rigidity. These compliance/stiffness maps can provide meaningful information for predicting the local and global resistance of the molecule against external pulling forces and, for example, they can be very useful when investigating the anisotropic response of proteins during mechanical unfolding experiments, at least for small pulling forces. Also, these maps can be used in the future to predict the packing of proteins.

The ANM is not the only model that has been developed for the prediction of protein flexibility. As a matter of fact, it has been shown that the Gaussian Network Model (GNM)³ often outperforms the ANM for B-factors predictions. Although the GNM does not consider the directionality of protein motion, the compliance and stiffness metrics can also be applied within the GNM with some modification. This is shown in the Supplementary Material, where the GNM compliance and stiffness maps and profiles have been derived and compared to B-factors and GNM cross-correlations maps. From the results, we observe that the GNM-based compliance and stiffness maps are, of course, less informative compared to the ANM-based ones, since they lack the information about the directionality of force application. Moreover, it is found that the GNM compliance profile basically overlaps with the fluctuations, thus leading to the same correlation

coefficient with the experimental data. Therefore, although the GNM fluctuations are often found to yield better correlations with B-factors compared to ANM ones, the application of compliance and stiffness concepts within the GNM seems less suitable, compared to the ANM, due to the lack of the directionality in the GNM.

In this paper, only the isotropic experimental B-factors have been considered for comparison with the fluctuations, compliances and stiffness profiles. Due to the directionality element within the suggested force application method, pairwise compliances and stiffnesses might be developed to provide also a metric of the anisotropic B-factors, which would be one further advancement of this research. However, to this purpose, several issues would need to be first addressed. First, the ANM-based fluctuations have already been compared to a limited extent to the anisotropic B-factors and the correlations were not found as good as in the comparison against the isotropic counterpart¹⁰. Second, the suggested force application methodology is strongly dependent on the protein geometry, so that residues on specific part of the molecule, e.g. on the surface, may experience a more limited directionality pattern in the force application compared to other residues. Also, although the pairwise force application seems the most rational choice to simulate the intra-molecular interactions among residues, the effect of external solvent is believed to be more random in nature. For all these reasons, the model and methodology presented here should be improved before making anisotropic B-factors prediction. This might be achieved: by considering explicitly solvent molecules embedded within the protein ENM, e.g. by taking also into account the layer of tightly bound water, ions, and small molecules; by considering a more random force application pattern, that allows accounting for both intra-molecular interactions and the bombardment due to external solvent; by considering an enhanced ENM, e.g. with more accurate spring connectivity and force constants. All these elements might be beneficial for the improvement of the model and the prediction of protein anisotropic B-factors.

Eventually, it should be also observed that B-factors may not be the best benchmarks, as some contribution may also arise from rigid-body motions³⁵. It was also shown that models that provide better correlations with crystallographic B-factors might still model collective motions less reliably³⁶. All these aspects should therefore be taken into account.

Finally, it is remarkable that the concepts of compliance and stiffness, which are used in the field of engineering to assess civil and mechanical structures and to analyze their structural responses, have found application to the field of protein dynamics. As an example, some of the authors recently developed a stiffness-based methodology based on matrix calculus, for the investigation of the structural behavior of special types of tall buildings, called diagrid, which are made up of diagonal members placed all over the exterior of the building³⁷. The matrix-based method proposed for the diagrid systems is found to share many mutual features with the ANM, which has originally been developed for the investigation of protein vibrations. It is thus fascinating that such different systems having extremely different distance scales, such as proteins and tall buildings, can be effectively investigated by the same structural approaches.

References

- [1] Tirion MM. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys Rev Lett* 1996;77:1905-1908.
- [2] Atilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, Bahar I. Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys J* 2001;80:505-515.

- [3] Bahar I, Atilgan AR, Erman B. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold Des* 1997;2:173-181.
- [4] Doruker P, Jernigan RL. Functional motions can be extracted from on-lattice construction of protein structures. *Proteins* 2003;53:174-81.
- [5] Doruker P, Jernigan RL, Bahar I. Dynamics of large proteins through hierarchical levels of coarse-grained structures. *J Comput Chem* 2002;23:119-27.
- [6] Sen TZ, Feng Y, Garcia JV, Kloczkowski A, Jernigan RL. The extent of cooperativity of protein motions observed with elastic network models is similar for atomic and coarser-grained models. *J Chem Theory Comput* 2006;2:696-704.
- [7] Tama F, Sanejouand YH. Conformational change of proteins arising from normal mode calculations. *Protein Eng* 2001;14:1-6.
- [8] Mahajan S, Sanejouand YH. On the relationship between low-frequency normal modes and the large-scale conformational changes of proteins. *Arch Biochem Biophys* 2015;567:59-65.
- [9] Eyal E, Yang LW, Bahar I. Anisotropic network model: systematic evaluation and a new web interface. *Bioinformatics* 2006;22:2619-2627.
- [10] Yang L, Song G, Jernigan RL. Protein elastic network models and the ranges of cooperativity. *Proc Natl Acad Sci U S A* 2009;106:12347-12352.
- [11] Tama F, Gadea FX, Marques O, Sanejouand YH. Building-block approach for determining low-frequency normal modes of macromolecules. *Proteins* 2000;41:1-7.
- [12] Hoffmann A, Grudinin S. NOLB: nonlinear rigid block normal-mode analysis method. *J Chem Theory Comput* 2017;13:2123-2134.
- [13] Scaramozzino D, Lacidogna G, Piana G, Carpinteri A. A finite-element-based coarse-grained model for global protein vibration. *Meccanica* 2019;54:1927-1940.
- [14] Kim MK, Jernigan RL, Chirikjian GS. Efficient generation of feasible pathways for protein conformational transitions. *Biophys J* 2002;83:1620-1630.
- [15] Atilgan C, Atilgan AR. Perturbation-response scanning reveals ligand entry-exit mechanisms of ferric binding protein. *PLoS Comput Biol* 2009;5:e1000544.
- [16] Atilgan C, Gerek ZN, Ozkan SB, Atilgan AR. Manipulation of conformational change in proteins by single-residue perturbations. *Biophys J* 2010;99:933-943.
- [17] Liu J, Sankar K, Wang Y, Jia K, Jernigan RL. Directional force originating from ATP hydrolysis drives the GroEL conformational change. *Biophys J* 2017;112:1561-1570.
- [18] Dykeman EC, Sankey OF. Normal mode analysis and applications in biological physics. *J Phys Condens Matter* 2010;22:423202.
- [19] Bermann HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The protein data bank. *Nucleic Acids Res* 2000;28:235-242.
- [20] Yang L, Song G, Carriquiry A, Jernigan RL. Close correspondence between the motions from principal component analysis of multiple HIV-1 protease structures and elastic network modes. *Structure* 2008;16:321-330.
- [21] Yang LW, Eyal E, Chennubhotla C, Jee J, Gronenborn AM, Bahar I. Insights into equilibrium dynamics of proteins from comparison of NMR and X-ray data with computational predictions. *Structure* 2007;15:741-749.
- [22] Khade PM, Kumar A, Jernigan RL. Characterizing and predicting protein hinges for mechanistic insight. *J Mol Biol* 2019;432:508-522.
- [23] Arikawa K. Theoretical framework for analyzing structural compliance properties of proteins. *Biophys Physicobiol* 2018;15:58-74.

- [24] Lamm G, Szabo A. Langevin modes of macromolecules. *J Chem Phys* 1986;85:7334-7348.
- [25] Woolard EW, Einstein A, Furth R, Cowper AD. Investigations on the theory of the Brownian movement. *Am Math Mon* 1928;35:318-320.
- [26] Cohen AE, Moerner WE. Suppressing Brownian motion of individual biomolecules in solution. *Proc Natl Acad Sci U S A* 2006;103:4362-4365.
- [27] Di Rienzo C, Piazza V, Gratton E, Beltram F, Cardarelli F. Probing short-range protein Brownian motion in the cytoplasm of living cells. *Nat Commun* 2014;5:5891.
- [28] Hinsen K, Petrescu AJ, Dellerue S, Bellissent-Funel MC, Kneller GR. Harmonicity in slow protein dynamics. *Chem Phys* 2000;261:25-37.
- [29] Ho BK, Agard DA. Probing the flexibility of large conformational changes in protein structures through local perturbations. *PLoS Comput Biol* 2009;5:e1000343.
- [30] Eren D, Alakent B. Frequency response of a protein to local conformational perturbations. *PLoS Comput Biol* 2013;9:e1003238.
- [31] Eyal E, Bahar I. Toward a molecular understanding of the anisotropic response of proteins to external forces: insights from elastic network models. *Biophys J* 2008;94:3424-3435.
- [32] Mishra SK, Jernigan RL. Protein dynamic communities from elastic network models align closely to the communities defined by molecular dynamics. *PLoS Comput Biol* 2018;13:e0199225.
- [33] McClendon CL, Kornev AP, Gilson MK, Taylor SS. Dynamic architecture of a protein kinase. *Proc Natl Acad Sci U S A* 2014;111:4623-4631.
- [34] Chopra N, Wales TE, Joseph RE, Boyken SE, Engen JR, Jernigan RL, Andreotti AH. Dynamic allostery mediated by a conserved tryptophan in the tec family kinases. *PLoS Comput Biol* 2016;12:e1004826.
- [35] Dehouck Y, Bastolla U. The maximum penalty criterion for ridge regression: application to the calibration of the force constant in elastic network models. *Integr Biol* 2017;9:627-641.
- [36] Fuglebakk E, Reuter N, Hinsen K. Evaluation of protein elastic network models based on an analysis of collective motions. *J Chem Theory Comput* 2013;9:5618-5628.
- [37] Lacidogna G, Scaramozzino D, Carpinteri A. A matrix-based method for the structural analysis of diagrid systems. *Eng Struct* 2019;193:340-352.

Table 1. Datasets characteristics

Dataset	1	2
Number of chains [-]	Single (1)	Multiple (2-10)
Resolution [\AA]	0.0 – 1.3	0.0 – 1.1
Sequence identity [%]	30	50
Number of structures [-]	921	149
Protein size (number of residues) [min – max]	101 – 1174	104 – 2484

Figure Legends

Fig. 1. **Calculation of structural compliance.** (a) Compliance for a linear spring with force constant γ . The spring is initially in the undeformed configuration (upper) and is then subjected to two equal and opposite forces F , resulting in a total elongation δ (lower). The compliance of the system is defined as the total resulting displacement in the direction of the force, divided by the value of the force, i.e. $C = \delta/F$. (b) Compliance related to residues i and j within the protein elastic network. A couplet of opposite unitary forces are applied at residues i and j , in the i - j direction, and the corresponding displacements δ_i and δ_j are evaluated. Note that the displacements δ_i and δ_j will not necessarily be parallel to the direction of the applied forces. Pairwise compliance is then evaluated according to Eq. (12).

Fig. 2. **Comparison of the fluctuation, compliance and stiffness correlations with the experimental B-factors for Dataset 1 ($p = 3$).** (a) Distribution of the correlation coefficients for fluctuations (grey histogram) and compliance (black histogram). (b) Distribution of the correlation coefficients for stiffness. Median values (M) and standard deviations (σ) are reported in the keys.

Fig. 3. **Comparison of the fluctuation, compliance and stiffness correlations with the experimental B-factors for Human Complement Protein C8 γ (PDB: 1LF7).** In (a) fluctuations; (b) compliances and stiffnesses; (c) graphical versions of these values shown on the structure – left, compliance; center, normalized B-factors; and right, normalized fluctuations. Coloring is spectral with red for higher values and dark blue for lower values.

Fig. 4. **Comparison of the fluctuation, compliance and stiffness correlations with the experimental B-factors for Dataset 2 ($p = 3$).** (a) Distribution of the correlation coefficients for fluctuations and compliance. (b) Distribution of the correlation coefficients for stiffness. Median values (M) and standard deviations (σ) are reported in the keys at the top.

Fig. 5. **Comparison of the fluctuation, compliance and stiffness correlations with the experimental B-factors for *Clitocybe nebularis* ricin B-like lectin (PDB: 3NBC).** In (a) fluctuations; (b) compliances and stiffnesses; (c) graphical versions of these values shown on the structure – left, compliance; center, normalized B-factors; and right, normalized fluctuations. Coloring is spectral with red for higher values and dark blue for lower values.

Fig. 6. **Distribution of the correlation coefficient differences Δ ($p = 3$).** (a) Dataset 1; (b) Dataset 2. Δ is defined as the difference between the correlation coefficient for compliance and the correlation coefficient for the fluctuation with respect to the experimental B-factors. Positive Δ values mean that the compliance has a better agreement with experimental data; negative Δ values imply a better agreement for the ANM fluctuations. For both Dataset 1 and Dataset 2, the distribution of Δ values is almost centered at zero, with a slight bias towards positive values.

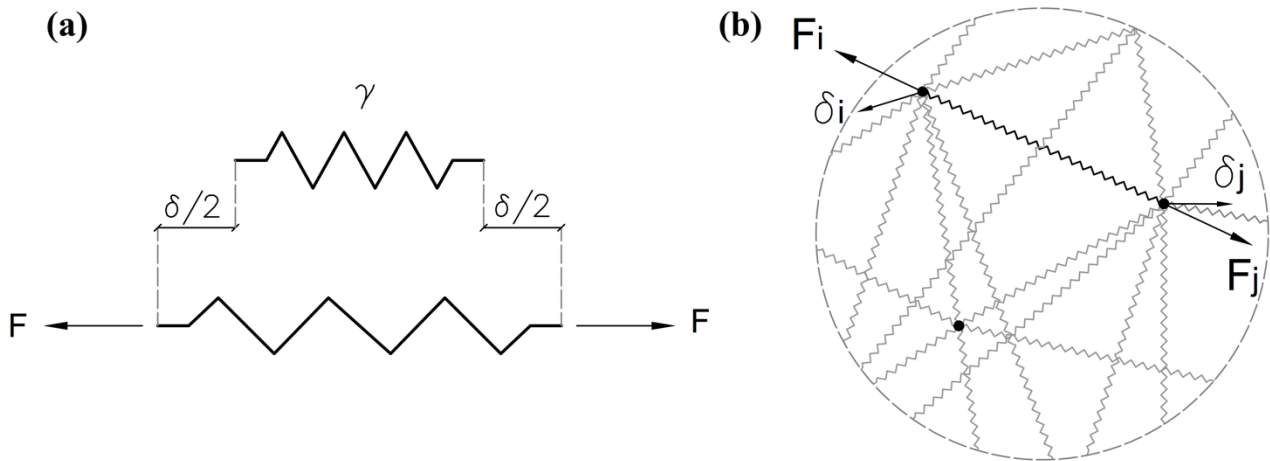


Fig. 1. Calculation of structural compliance. (a) Compliance for a linear spring with force constant γ . The spring is initially in the undeformed configuration (upper) and is then subjected to two equal and opposite forces F , resulting in a total elongation δ (lower). The compliance of the system is defined as the total resulting displacement in the direction of the force, divided by the value of the force, i.e. $C = \delta/F$. (b) Compliance related to residues i and j within the protein elastic network. A couplet of opposite unitary forces are applied at residues i and j , in the i - j direction, and the corresponding displacements δ_i and δ_j are evaluated. Note that the displacements δ_i and δ_j will not necessarily be parallel to the direction of the applied forces. Pairwise compliance is then evaluated according to Eq. (12).

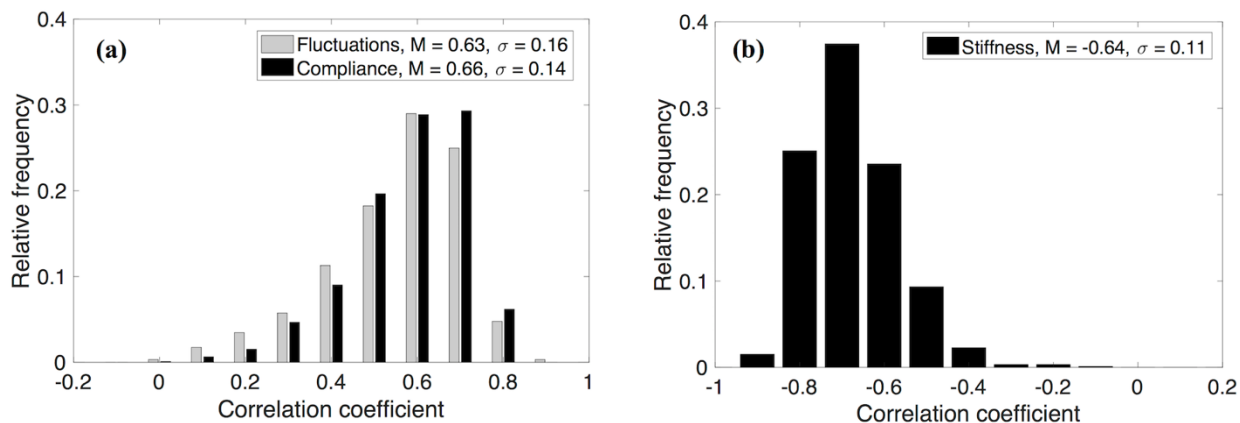


Fig. 2. **Comparison of the fluctuation, compliance and stiffness correlations with the experimental B-factors for Dataset 1 ($p = 3$).** (a) Distribution of the correlation coefficients for fluctuations (grey histogram) and compliance (black histogram). (b) Distribution of the correlation coefficients for stiffness. Median values (M) and standard deviations (σ) are reported in the keys.

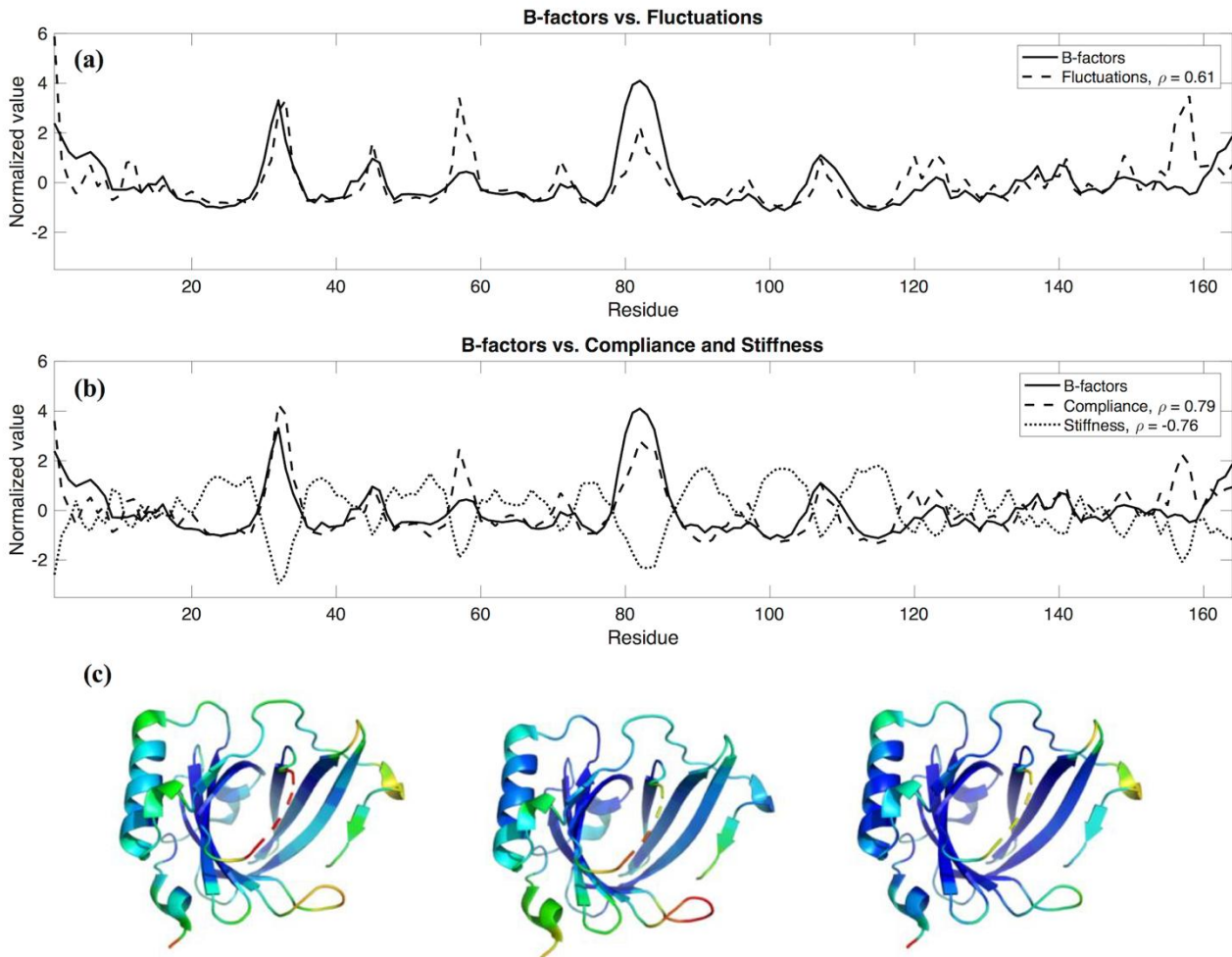


Fig. 3. **Comparison of the fluctuation, compliance and stiffness correlations with the experimental B-factors for Human Complement Protein C8 γ (PDB: 1LF7).** In (a) fluctuations; (b) compliances and stiffnesses; (c) graphical versions of these values shown on the structure – left, compliance; center, normalized B-factors; and right, normalized fluctuations. Coloring is spectral with red for higher values and dark blue for lower values.

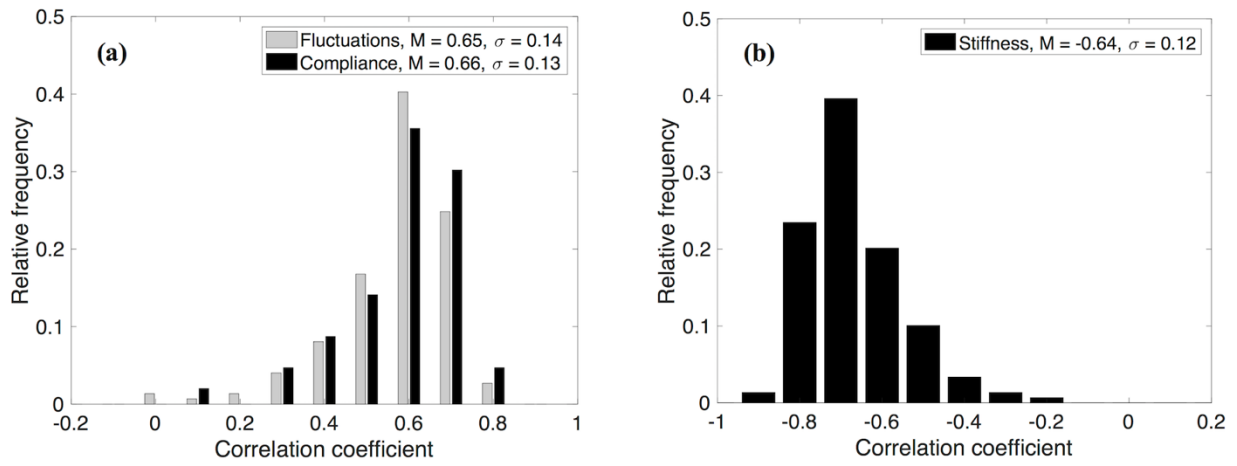


Fig. 4. Comparison of the fluctuation, compliance and stiffness correlations with the experimental B-factors for Dataset 2 ($p = 3$). (a) Distribution of the correlation coefficients for fluctuations and compliance. (b) Distribution of the correlation coefficients for stiffness. Median values (M) and standard deviations (σ) are reported in the keys at the top.

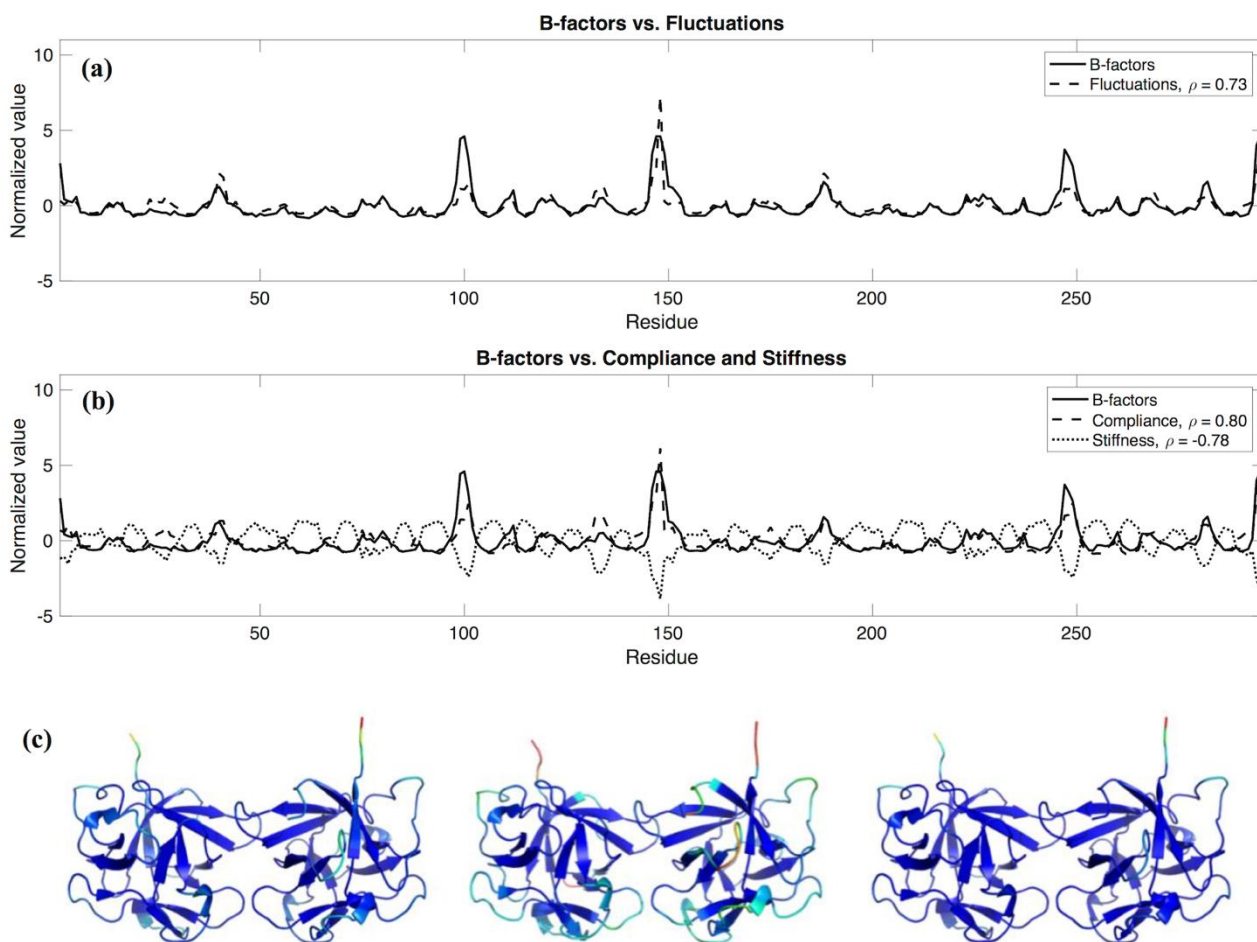


Fig. 5. Comparison of the fluctuation, compliance and stiffness correlations with the experimental B-factors for *Clitocybe nebularis* ricin B-like lectin (PDB: 3NBC). In (a) fluctuations; (b) compliances and stiffnesses; (c) graphical versions of these values shown on the structure – left, compliance; center, normalized B-factors; and right, normalized fluctuations. Coloring is spectral with red for higher values and dark blue for lower values.

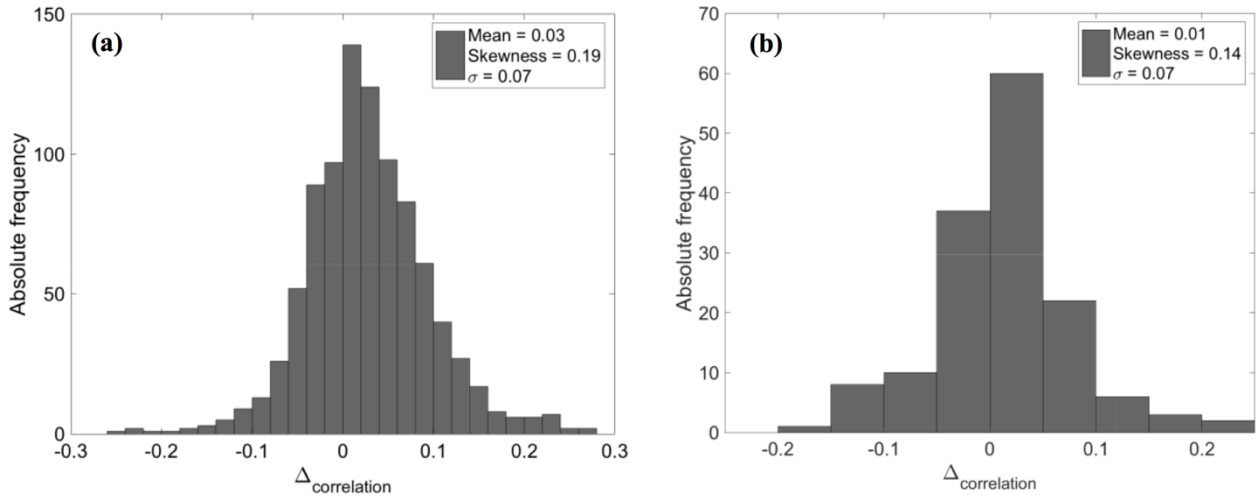


Fig. 6. **Distribution of the correlation coefficient differences Δ ($p = 3$).** (a) Dataset 1; (b) Dataset 2. Δ is defined as the difference between the correlation coefficient for compliance and the correlation coefficient for the fluctuation with respect to the experimental B-factors. Positive Δ values mean that the compliance has a better agreement with experimental data; negative Δ values imply a better agreement for the ANM fluctuations. For both Dataset 1 and Dataset 2, the distribution of Δ values is almost centered at zero, with a slight bias towards positive values.